

DOCUMENT RESUME

ED 410 495

CG 027 969

AUTHOR Kreft, Ita G. G.
TITLE The Evaluation of Drug Prevention Programs and Its Effect on Alcohol Consumption of Junior High School Students.
SPONS AGENCY National Inst. on Drug Abuse (DHHS/PHS), Rockville, MD.
PUB DATE 1997-03-00
NOTE 35p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).
CONTRACT DA09649-02
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Adolescents; *Drinking; *Drug Education; Evaluation Needs; Evaluation Problems; *Junior High School Students; Junior High Schools; Predictor Variables; Prevention; Program Effectiveness; Statistical Bias; Student Behavior
IDENTIFIERS Drug Abuse Resistance Education Program

ABSTRACT

Many reports of successful school-based intervention programs suffer from an overestimation of the real effects, errors which many times arise from statistical technicalities. An illustration of such overestimations, as seen through a reanalysis of data on a successful program, is reported in this paper. The data were reevaluated using a variety of statistical techniques, such as multilevel analysis. The result of the analyses shows that the reported effects of the normative school-based drug prevention program could not be replicated. The subsequent search for moderator effects of the program, such as a lowering effect on the relationship between the pre- and post-test, or on the relationship between respondents' use and their friends' use, was not successful either. More successful was a search for individual characteristics that show significant relationships with respondents' alcohol use, such as school problems, low grades, and rebelliousness. (RJM)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED 410 495

**The Evaluation of Drug Prevention Programs and
its effect on Alcohol Consumption of Junior Highschool
Students**

Ita G.G. Kreft
California State University, Los Angeles
Charter School of Education

This research was supported by NIDA grant #DA09649-02
Nida grant # DA09649-02 provided me with six month over a period of two years (1994 -
1996) to reanalyze the Adolescent Alcohol Prevention Trial data, collected by Hansen and
Graham. The data is collected and made available by John W. Graham

PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

I. Kreft

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- ☐ This document has been reproduced as
received from the person or organization
originating it.
- ☐ Minor changes have been made to improve
reproduction quality.

* Points of view or opinions stated in this docu-
ment do not necessarily represent official
OERI position or policy.

CG027969

BEST COPY AVAILABLE

ABSTRACT

Many reports of successful school based intervention program suffer from an overestimation of the real effect, due to statistical technicalities, such as the choice of the unit of analysis and measurement errors. This paper is an illustration of this overestimation by reanalyzing data reported about a successful program, Normative Education. With the support of a NIDA grant the data is thoroughly taken apart, using different statistical techniques. The result of the analyses show that the reported effects of the normative school based drug prevention program could not be replicated. The subsequent search for moderator effects of the program, such as a lowering effect on the relationship between the pre- and post-test, or on the relationship between respondents use and the use of their friends, was not successful either. For this data I have reasons to believe that the null hypothesis of zero effects of the program should be retained, since reported analyses results of the same data are inconsistent. More successful was a search for individual characteristics that show significant relationships with respondents' alcohol use. Among them was the abuse of alcohol by adults in respondents' direct social environment.

INTRODUCTION

School based drug prevention programs have been a part of the US 'war on drugs' campaign over the past twenty years. The most widely used program among them is D.A.R.E.ⁱ. Studies that evaluate the effects of D.A.R.E. are disappointing. Consequently alternative programs have been developed that try to avoid some real or imagined flaws of D.A.R.E.ⁱⁱⁱ.

In a review of the literature I found studies that went beyond the general question: "Do school based drug prevention programs work or not"? The studies can be divided into two groups, qualitative, where theory is grounded in empirical data, and quantitative, where a theory is applied to make changes in existing programs. The qualitative approach is discussed in the paper of D'Emidio-Caston and Brown, in this issue. Examples of quantitative approaches, where theory guides the data collection, are found in Graham, Marks, and Hansen (1991, see also Hansen, 1993, and May, 1993), where the components of resistance training programs (which is a D.A.R.E. concept) are defined and compared with social theories (e.g. Bandura, 1977 and 1986, and Jessor & Jessor). Hansen, Graham, Bolkenstein and Rohrbach's 1991 analyses show that resistance training significantly improved adolescent refusal skills, but the same skills failed to predict less alcohol use. They propose a new program, based on social theory, called "normative education", which seem to be successful in lowering the onset of alcohol use in teenagers (Hansen and Graham, 1991).

The conclusion that resistance training alone does not work is reported in other literature as well. An overview by Dukes, Ullman and Stein (1996) concludes that the effectiveness of D.A.R.E. is inconsistent, where most studies indicate very little or no effects. Ennett, Tobler, Ringwalt and Flewelling (1994) reach the same conclusion. In a meta analysis of short term effects of D.A.R.E. they find slight and, except for tobacco use, not statistically significant effects (l.c. p.1398).

The inconsistencies in the reported success of drug prevention programs may be partly due to the methods used to analyze the data. Data are sometimes analyzed at an aggregated class level (see Duke, Ullman and Stein, 1995, Hansen and Graham, 1991), while the results are presented as being valid for individual students, resulting in ecological fallacies (see on this topic Robinson 1950, Kreft and DeLeeuw, 1988, among others). In some analyses the response variable is dichotomized, where alcohol use is coded as one, and non-use as zero (e.g. Hansen and Graham, 1991, and Graham, Collins, Wugalter, Chung, and Hansen, 1991). In such cases the distinction between abuse, use, or just one single glass of alcohol disappears. Since drug prevention programs deal largely with teenagers, who are experimenting with life, a single experiment will put them in the users/abusers category. This type of data management may hide more than it reveals.

Another issue is the use and abuse of statistical testing. Evaluation of drug prevention programs is sometimes based on a single statistical test of significance. There exists a vast literature about the merits of statistical testing, and the interpretation and misinterpretation of such tests (e.g. Wang, 1993). A mistake easily made is to think that a large statistical significance between the control group and the program is an indication of large numbers of students saved. But a small p-value or a large t-statistic is only a confidence statement regarding the rejection of the null hypothesis. If drug prevention evaluations are based on many observations, as they mostly are, a small difference in numbers of students abstaining from alcohol between the program and the control group can result in a significant statistical difference.

More important information, about the strength of the significant effect, is missing in many of reports of successful programs. The decrease in numbers of alcohol users in a successful program, as compared to other conditions, are hidden behind F- or T-tests. As in my analysis of abstainers, reported later in this paper, the success of a normative program is based on a difference of nine students compared to the control group. And the 62.4% reduction in the rate of onset of drunkenness attributed to the normative program, reported in Hansen and Graham (1991, p.422-423), is only 32 students, since drunkenness is relative rare. It is hard, but possible, to manipulate this data until this difference shows up in a statistically significant cross table ($P=0.04$ Pearson Chi-square for the total table).

The abuse of statistical testing is sometimes hidden, when significant results are reported.

without mentioning the number of non-significant results. Based on statistical theory we know that one out of twenty tests will show significant effects by chance alone. The few effects reported in the literature may suffer from this capitalization on chance, since in this type of research a choice among many response variable can be made. In the data I reanalyzed for instance, there are twelve variables that relate to actual drinking and intentions to start drinking. All of them are potential response variables.

The twelve response variables present in the data of Hansen and Graham (1991) are used in this paper to create a single summarized response variable, labeled alcohol involvement of respondents. By using a scaling method for categorical data a continuous scale is obtained, where high use is scaled low (negative) and no use is scaled high (positive). In all the different analyses this student level variable is used as the response variable.

The report of my re-analysis of the Hansen and Graham data consists of two distinctive parts with two purposes. In the first part analyses techniques are introduced with the purpose of elucidating the statistical problems and fallacies mentioned earlier. Two suitable techniques became available recently in user-friendly software packages. One is especially developed for analyzing hierarchically nested data, and allows to analyze the data at class and student level (ML3, 1989). The other technique is useful for data reduction and the transformation of ordinal scaled items into a single continuous variable (HOMALS, 1989). Using this last technique generally results in a scale with less measurement error compared to the original items.

In the second part the actual analyses are executed where the effects of drug prevention programs are evaluated. The case for zero effects of programs is made as strong as possible, by using different analyses methods, a technique known as triangulation. Triangulation can show if weak effects are merely happening by chance alone. If an effect is weak, it may show up in one analysis, but not in another, which is an indication of a questionable result. All of this ties in with the earlier discussion: if an effect is statistically significant, how large is that effect, as expressed in interpretable numbers?

DESCRIPTION OF THE DATA

The data are collected by the Adolescent Alcohol Prevention Trial, a longitudinal drug prevention trial, examining two psychologically based strategies for preventing the onset of adolescent alcohol and drug use (AAPT), a longitudinal study, that measures students over several years. For this paper the pre-test measurement of 1987 in seventh grade of junior high school is

used and the measurement one year later, after the implementation of the treatment, in eighth grade. The twelve schools in the study consist of 118 school classes. Seven of the twelve schools are public schools, while five are Catholic schools. Schools are randomly assigned to one of four experimental conditions, which are:

- (1) a control group of 32 classes receiving the general information program about consequences of use only, (CONTROL)
- (2) a group of 33 classes receiving resistance training (RT) + information,
- (3) Normative education (NORM^v) + information was received by 27 classes, and
- (4) a group of 26 classes receiving a combined condition, with RT + NORM + information (BOTH) (see Hansen and Graham, 1991, for more detailed information).

The pre-test sample consists of 3027 students, and the post-test data contains the answers of 3147 respondents, 2378 students answered both questionnaires, one in 1987 and in 1988. For analyses purposes two levels are recognized, the students as the first level, and classes as the second level of observation. Classes are an equally important level of analyses as students are for this study. A class climate can be one of the influences that make drug prevention programs a success, or a failure, and for that reason should be ignored in the evaluation of such programs. Information at both levels is present in the data. The most important class level measurement is the drug prevention program, which has four different categories, as described above. The programs are dummy coded, with the control group as reference. At student level over two hundred items are present. For the analyses in this paper thirty three are picked to construct four new variables. The four variables are: pre-test for alcohol involvement of the respondent, the respondents' post-test (which is the response variable), the alcohol use of friends, and the alcohol use by adults in the respondents' environment. Several of these student variables are averaged and used as an indication of class climate.

The most important goal of the analyses is to find effects of drug prevention programs, and especially of the reported successful program NORM. It is known that programs can have general effects on students' behavior, but also moderator effects. Moderator variables are defined in regression analyses as interacting variables, interacting with the relationship between explanatory and response variables. To investigate if moderator effects are present in this data, a new type of variable are created, known in the multilevel literature as crosslevel interactions. The name 'crosslevel' indicates that the interaction involves variables from different levels of the hierarchy, here the student level and the class level. By multiplying two variables, one from each level, crosslevel interaction terms are constructed for the purpose of investigation. The program NORM is used to create two interactions by multiplying this variable with two student level variables, the respondents' pre-test and friends' alcohol use. If a significant moderator effect of NORM is present in the data it will be an indication that this program has an effect on the relationship of the

interacting variable and the response variable. When that effect is significant, and negative, the conclusion can be drawn that NORM has a moderator effect by lowering the relationship between pre- and post-test, or between friends' alcohol consumption and respondents' post-test alcohol involvement.

The surveys are collected one year apart, first in seventh grade, and one year later, when the same students are in eighth grade. In both years the same questionnaire is applied, containing many items that measure alcohol use of respondent, use by friends and use by adults in the social environment of the respondents. The two data sets combined (the one measured in 1987 and the one in 1988), resulted in a data set with a total of 2378. Due to missing cases in one or the other year, 649 students had to be deleted.

DESCRIPTION OF ANALYSES TECHNIQUES

The paper has two quite distinct objectives. The most obvious one is the evaluation of drug prevention programs, the other is the introduction to fairly recent developed techniques, multilevel and homogeneity analysis. The techniques are briefly introduced in the next paragraphs, since it is expected that they are relatively unknown to most readers. For a more extensive description, as well as an explanation of the new modeling concepts behind these techniques, I refer to the relevant literature. For details of applications to drug prevention research I refer to Kreft (1997) and Kreft (1994).

Multilevel Analysis

Multilevel analysis is especially developed for data collected in situations where observations are clustered in groups. In the evaluation of drug prevention programs observations are mostly clustered within programs, while programs are administered to existing groups, such as school classes, health groups, or community centers. The goal of drug prevention programs is to influence individual behavior, while randomization and implementation of the intervention occurs at the group level. In our data the individuals are junior high school students in seventh and eighth grade, clustered in 118 school classes. Prevention programs are applied at the class level. Students and classes are both important levels of observation. First existing classes are not equal to randomized groups, but have over time developed their own class climate, and class dynamics. As a result of this interaction among classmates, students in the same class are more alike than students in different classes, resulting in dependent observations. Also, students in the same class share behaviors and value systems that may interfere with the failure or success of prevention programs. In multilevel analyses both levels are defined as levels of influence, and both exert

influence on the behavior of students. In this data even more levels are present, the first level is the individual student, nested within the second level, the class, while the class is nested in a third level, the school. For the analyses in this paper only two of the three levels are considered, the student and the class. The class level is chosen over the school level because of its importance in relation to drug prevention programs. The school level could be included as a third level, but since only 12 schools are present in this data, the school level is ignored. The two levels analyzed in the multilevel analyses in this paper are students (level 1) and classes (level 2).

Before any analyses techniques were available to analyze both levels at the same time discussions centered around the question: what is the appropriate level of statistical analysis for this type of data, student level or class level? (Robinson, 1950, Hannan, 1974, Burstein, 1980, Kreft and De Leeuw, 1988). The conclusion reached in the literature of that time is that class level analyses can deliver the wrong message in relation to student behavior, since it can only make conclusions about classes. Inferences made about individuals based on aggregated data can be a lead to wrong conclusions, as illustrated almost half a century ago by Robinson (1950), who labeled this the 'ecological' fallacy. Robinson's examples show that aggregated correlations can have an opposite sign compared to individual correlations, even when calculated over the same data. For the same effect on regression coefficients see Kreft and De Leeuw (1988), who labeled this effect the 'see-saw' effect.

Multilevel analyses does take that intraclass correlation into account, and analyses can be done at student level, while also analyzing the data at class level. One of its main advantages is that many more research questions can be answered, as compared to single level analyses. It allows to explore the data at both levels, and to discover the complex relations of variables measured at different levels with the alcohol use of respondents. Research questions that mix levels of influence together are common in multilevel analyses. For example the question, "What is the effect of class climate, together with the effect of alcohol consumption of friends, on respondent's alcohol use?" can be answered. Aggregation of student level measurements to the class level deletes the important variation in respondents' personal relationships with family and friends. In research where important variation in social environment of respondents is deleted, 'at risk' students can no longer be identified. The conclusion is that student level variation cannot be ignored, but neither can class variation. Drug prevention programs are administered to classes, and class climate may interfere with the success of any of these programs. In the multilevel analyses reported later, both levels are taken into account. Questions related to students' individual behavior, as well as the behavior of school classes, are asked and answered.

History of Multilevel Models

Multilevel regression models have been developed for analysis of hierarchically nested data,

such as students nested in classes. These techniques correct for intraclass correlation, and allow the researcher to estimate individual effects, as well as group-level and drug prevention program effects. In the models it is possible to test cross-level interactions, such as between drug prevention program conditions and student characteristics. Such effects are also known as moderator effects, as discussed earlier.

Multilevel methods are presently applied in many different fields, such as education (Aitkin and Longford, 1986, De Leeuw and Kreft, 1986, Raudenbush and Bryk, 1987), public health (Laird and Ware, 1982, Longford, 1987), and psychiatry (Hedeker et al. 1991, 1995). At the moment many programs are available for this type of analyses (see Kreft, De Leeuw and Van der Leeden, 1994 for a review of some of these programs). Carbonari, Wirtz, Muenz and Stout (1994, p.89) consider multilevel methods "an important advance in the field." They specifically mention as a virtue of the new model the resolution of the issue of the unit of analysis for unbalanced data.

For this paper the software ML3 (1989) is used.

Scaling variables with homogeneity analysis

The second technique described is Homogeneity analysis, a data reduction technique that scales several variables into a single continuous variable. Four scales in total will be constructed, where two of the most important ones are respondents alcohol involvement in pre- and post-test. Both are measured with twelve variables. In 1987, and again in 1988, twelve questions are asked related to present and anticipated alcohol use. All twelve are used to construct one pre- and one post-test measurement, representing respondents' alcohol involvement in both years. The twelve questions are:

- Item 19: How many drinks of alcohol have you had in your whole life? 1=none, 9= > 100
- Item 20: How many drinks in the past month? 1=none, 8=> 20
- Item 23: How many drinks in the past week? 1=none, 8=11 or more
- Item 24: How many days in the past month did you have a drink? 1=none, 6= 15 to 30
- Item 25: How long since you had a drink of alcohol? 1= less than 24 hours, 7= never a drink
- Item 28: Think of the day during the past month when you drank the most alcohol. How many drinks did you have on that day? 1=I never drink, 8=5 or more
- Item 29: How often do you imagine having a drink? 1=often, 4= never
- Item 32: Do you think you will drink in the next months? 1= yes, 4= no.
- Item 33: Do you think you will ever drink alcohol every day? 1= yes, 4= no
- Item 34: Do you think you will ever drink every month? 1= no, 4= yes
- Item 35: How many times have you ever been drunk? 1= never, 6= > than 20 times
- Item 38: Do you think you will get drunk in the next month? 1= no, 2= yes

(For more details, answer categories and frequencies see the appendix)

Homogeneity analysis is a method developed for the analyses of categorical variables, as most variables in the list above are. The technique is used in my example to construct a numerical scale based on all items present. The software used is HOMALS, which is an acronym for HOMogeneity analysis by Alternating Least Squares, and available in Categories (SPSS, 1989). This type of analysis is comparable to principal component analysis (PCA), but for categorical instead of for numerical data. The method has been given many names, because it was discovered independently by different people, but first applied by Guttman (1941) for the scaling of constructs. The best known name for this technique is multiple correspondence analysis, used by Benzecri (1973) and Greenacre (1984). Other names are dual scaling, method of reciprocal averaging, linearization of regression, and seriation (see Van de Geer, 1993a, 1993b). If all variables are binary, results of HOMALS will be the same as those obtained from classical PCA.

The technique is first demonstrated using the variable for alcohol involvement of respondents, in pre-test and in the post-test year. As illustrated in Table 1, items measuring the same behavior are not always in agreement. The table shows that two items, measuring the same behavior, contain measurement error. Some of the respondents give answers that do not agree with previous answers given.

TABLE 1. A comparison of two items measuring alcohol use of respondents.
rows: Items 20: How many drinks of alcohol have you had in the past month?
columns: Item 24: How many days in the past month have you had alcohol to drink?

Categories item 24	none	1	2 or 3	4 to 7	8 to 14	15 to 30
Categories item 20						
1. none	2347	28	1	4		
2. only a sip (for religious service)	116	8	6			
3. only sips	89	104	31	5	1	1
4. part of a drink	17	56	23	3		
5. 2 to 4	13	36	44	16	3	3
6. 5 to 10	4	4	6	15	5	1
7. 11 to 20		2	3	3	4	1
8. more than 100		1	3	2	1	6

Note: The inconsistent answers are in bold face.

Many inconsistencies are found in Table 1, where 33 students (see the bold face numbers

28+1+4 in the first row) report not drinking last month, yet answer for item 24 that they had a drink at one or more days of that month. The reverse is also present. Of the students that answer to item 20 that they did not drink that month, except may be a sip, 85 students answer to item 24 that they had a drink at several occasions that month, where answers range from one to fifteen and more (see the bold face numbers in the first three rows, $28 + 1 + 4 + 8 + 6 + 31 + 5 + 1 + 1$). More inconsistencies are present, all illustrating that items measuring the same behavior are not always answered in the same way. The HOMALS construct for 'alcohol involvement' of respondents is based on the total answering pattern over all items, which takes inconsistencies in answering patterns into account. As a result a more reliable scale for alcohol involvement becomes available.

Scaling the variables with homogeneity analysis is also useful for other purposes. The technique can deal with missing data, preventing listwise deletion of cases when respondents have one or more answers missing. The missing data are replaced by values that are close to the values for students with similar, but complete answering patterns.

ANALYSES RESULTS OF SCALING

The items that measure pre- and post-test alcohol involvement are skewed to the right with a mode at category one, 'no alcohol use'. This skewness is present in most data on drug use in such a young population. The proportion of abstaining students on the items is between 55% and 95% (see appendix), depending on which of the twelve questions in the pre- or the post-test is observed. The lowest number of abstainers is found for the item that measures alcohol use over a life time (item 19), followed by the percentage of abstainers for drinking in the past months (item 20). The highest percentage of abstainers is found in the item measuring alcohol use in past week (item 23). Combining the twelve questions yield a single variable with a smaller proportion of abstainers than most of the separate items have, also resulting in a variable with better statistical properties.

Results of homogeneity analyses for scaling of respondents alcohol Pre-test and Post-test

In the first two analyses, reported in Table 2, the scales are constructed for the pre- and post-test, and labeled respondents' alcohol involvement. A conceptual difference among the items is indicated in the table by a line that divides the first eight items from the the last four items. The first items measure actual alcohol consumption, while the last measure drinking as projected in the future.

TABLE 2. Eigenvalues and discrimination measures for the years 1987 and 1988 smoking variables.

	Discrimination Measures*		Eigenvalues*		Number of observations	
	1987	1988	1987	1988	1987	1988
Item19	0.670	0.712	0.5566	0.6081	N=3027	N=3047
Item20	0.709	0.775				
Item23	0.550	0.580				
Item24	0.694	0.737				
Item25	0.660	0.694				
Item28	0.658	0.715				
Item35	0.516	0.562				
Item38	0.460	0.586				
Item29	0.421	0.498				
Item32	0.654	0.678				
Item33	0.232	0.249				
Item34	0.456	0.511				

* the correlation of each variable with the underlying scale

** a measure for the reliability of the scale

The table shows discrimination measures and eigenvalues. Discrimination measures are the factor loadings for each variable for each year, labeled discrimination measures. The eigenvalue for an analysis is a measure of overall fit, one for each year. Although the concept of discrimination measures is equal to the concept of factor loadings in traditional PCA, it would be misleading to use the same name, since the estimation methods are not comparable among methods. The discrimination measures in Table 2 show different values or loadings. These different values indicate the different contributions of variable to the underlying scale. Item 20 has the highest discrimination measure in both analyses, showing that this question (alcohol consumption over the last month) is the most important one.

The magnitude of the discrimination measures show if an item is an important contributor to the scale formed by all variables together. The items about weekly drinking (item 23) and about being drunk (item 35) contribute somewhat less to the scale, as indicated by the magnitude of the discrimination measure. The reason why these items do not discriminate equally strong as the other items is due to the high number of abstainers. The same is true for item 33 that measures the most extreme drinking behavior, "do you think you will ever drink alcohol every day". This question is answered by most student with 'no' (see the appendix) and resulting in a low discrimination measure. Items may not discriminate very highly, but are not deleted from the analysis. This decision is based on theoretical considerations, since all three items are strong in measuring alcohol involvement.

Discrimination measures can also be interpreted as the correlation with the underlying scale. In the two separate analyses in Table 2, item 20 has the highest correlation with the newly formed scale in 1987 as well as in 1988. Item 33 on the other hand has the lowest correlation in both years. The two analyses show similar patterns in the other variables as well, an indication of the reliability of the scales. A discrimination measure of a variable shows the proportion of variance of the variable that is between categories of that variable. Or, equivalently, a discrimination measure is the variance between the category quantifications of that variable. Consequently, a low value shows that the categories of that variable does not discriminate much.

The items that are the strongest determiners of present and future alcohol use are the items that measure monthly drinking (items 20, 24, 28 and 32), together with life time drinking (item 19). These items are also theoretically of the most importance, which makes the new constructed variable for pre- as well as for post-test a valid measure for our analyses of the evaluation of drug prevention programs.

Table 2 shows the eigenvalues for both analyses, which are 0.5566 and 0.6081 respectively. Eigenvalues are average discrimination measures, and can be used as an overall measure of fit. The highest possible value of eigenvalues (and discrimination measures as well) is 1.00. In the table the eigenvalue of 0.5566 for the analysis of items measured in 1987 indicates that 56% of the variation of the new scale is between categories of all variables. The items measured in the post-test explain 61% of the variation.

Category quantifications

The category quantifications of each item, part of HOMALS output, are reported in the appendix. It shows the new category values, which have a reverse ordering compared to the old category quantifications. The new quantifications indicate that high positive for pre- as well as for post-test indicates no alcohol use, while low negative indicates a high level of alcohol use. The same is true for the resulting scale, where a positive score means no drinking or positive behavior, and a negative score is alcohol use or abuse, or negative behavior. For analyses purposes the new category quantifications of the variables in the analyses are of no further use, but they are of interest for a better understanding of the new scale. The new scale as a mean of zero and a standard deviation of one.

A comparison between the HOMALS category quantifications with the original ones show that categories of an item are no longer equally distanced, the way the original categories are. For instance, item 23 (How many drinks in the past week) shows that the HOMALS distance between no drinking and 'a sip for religious purposes' is smaller and no longer equal to the distance between 'one sip' and 'one drink'. The new category quantifications behave very similar among the two analyses, an indication of the reliable of this scaling method as applied to this data.

In the next paragraphs variables are constructed that measure the alcohol use in the immediate environment of the respondents. One is friends' alcohol use, the second a construct for alcohol use by adults in the immediate environment of the respondents. Again HOMALS is used to summarize the available items.

Construction of a Pre- test for Friends' Alcohol Use

In the same way as before all variables that measure alcohol involvement of friends of the respondents are used to form a scale. Seven questions are available in the data about friends and their alcohol consumption and behavior in the pre-test 1987. These questions are,

- Item 1: How many of your three best friends have ever tried drinking alcohol?
- Item 2: How many of your best friends have had alcohol to drink in the past month?
- Item 2a: Has your best friends in your grade in this school had alcohol to drink in the past month?
- Item 3: How many of your best friends have ever been drunk?
- Item 4: How many of your three best friends have been drunk during the past month?
- Item 36: How often are you with kids who are drunk?

In Table 4, results are reported of the homogeneity analyses with the six available items measuring alcohol use of friends. The discrimination measures of the separate items show that items 2 and 3 are the most important contributors. They measure receptively drinking of friends in general and over the past month. The scale formed by the items is scaled in the same way as the pre- and the post-test, with a mean of zero and a standard deviation of one. The new variable is labeled 'friends alc' in the analyses reported later in this paper.

TABLE 4. Eigenvalues and discrimination measures for year 1987 and 1988 friends drinking alcohol.

	Discrimination Measures*	Eigenvalues**	Number of observations
Item 1	0.572	0.582	N=3027
Item 2	0.731		
Item 2a	0.565		
Item 3	0.634		
Item 4	0.570		
Item 36	0.420		

* the correlation of each variable with the underlying scale (similar to a component loading)

** a measure for the reliability for homogeneity of the scale

Construction of a scale for alcohol use in the immediate social environment

The next scaling is a construct based on three variables that measure drinking by adults in the environment of the respondent. For the construction of this scale three questions from the 1987 questionnaire are used, which are

- Item 26: How many times have you been offered a drink of alcohol in the past month?
- Item 30: How often are you with adults who are drinking alcohol?
- Item 37: How often are you with adults who are drunk?

The scale formed by the three variables is again scaled with high positive indicating low alcohol use, while high negative indicates high alcohol use. The name used for this variable in the analyses is 'social'.

TABLE 5. Eigenvalues and discrimination measures for year 1987 for drinking of alcohol in the environment.

	Discrimination Measures*	Eigenvalue**	Number of observations
Item 26	0.641	0.5457	N=3027
Item 30	0.312		
Item 37	0.684		

* the correlation of each variable with the underlying scale

** a measure for the reliability of the scale

In Table 5, the results of this homogeneity analysis are reported, which show that the middle question, item 30: "How often are you with adults who are drinking alcohol?" contributes the least to the scale. This supports the finding of Brown and Caston (1995) that there exists a difference between use and abuse. Drinking of alcohol can happen in a social context, and does not have to be abusive. Alcohol abuse is more indicated in the questions stated in items 37 and 26, which contribute equally strong to the scale, with discrimination measures of 0.641 and 0.684 respectively.

TRIANGULATION: SOME ANALYSES

Traditional Analyses: Multiple Regression and analysis of covariance

In the introduction I argued that analyzing these clustered data at one single level is a mistake, since one or the other level will be ignored. On the other hand, I expect that strong effects will even show up in 'flawed' methods, while weak effects may not always survive in different

analysis techniques. This method is known as triangulation. A comparison of effects based on different ways to analyze the data can support an earlier found effect, or it fails to support the earlier findings. If that happens the earlier findings are made questionable.

The first analysis technique presented in this paragraph is multiple regression with students as the unit of analysis. The second technique is analysis of covariance (ancova), where the programs are the factors, and student' alcohol use in 1988 is the response variable, while the alcohol use in 1977 is the covariate. The results obtained by these two traditional linear techniques are compared with the results obtained by a multilevel analysis.

All models in this section are simple. In the first model the pre-test predicts the post-test. In a next model the three drug prevention programs are added. The interpretation of the analyses results follow the same scenario. First I look at the value of the individual coefficients in the model and compare these with their standard errors. When an individual coefficient is significant it is indicated with a star. But more importantly, the total model fit is compared among different models. Decisions about significant effects are made based on the total fit of a model, rather than on significant tests of individual coefficients. Decisions based on individual coefficients may increase the chance of type 1 errors.

Multiple Regression

The first analyses are in Table 6, where a traditional multiple regression analysis is executed, using the two models described earlier.

Table 6. Multiple Regression between pre-test Homals-alcohol and post-test Homals-alcohol and two explanatory variables, Prevention Programs RT and NORM, N=2378.

Response Variable	alcohol88 (Homals) Model 1			Model 2		
	Coeff	St Error	R ²	Coeff	St Error	R ²
Constant	-0.03	0.02		-0.03	0.02	
alcohol87 (Homals)	0.65	0.02**	0.38	0.65	0.02**	0.38
NORM				0.11	0.05 *	
RT				-0.05	0.04	
BOTH				0.02	0.04	

* p < .05 ** p < .01

In the multiple regression analysis a statistically significant effect for NORM is present. This result is not supported by a better model fit, since the R² for both models (with and without programs) is equal. The explained variance in the both regression models is R² = 0.38.

Individual parameter interpretation would be capitalizing on chance, since the overall fit of the model does not show improvement.

Regression analyses is not the best model for the analysis of clustered data. Analysis of covariance, which treats the program conditions at the correct level, is a better approach. It still does not correct for intraclass correlation, but it can serve as a preliminary check for group effects before executing a multilevel analysis.

Analysis of Covariance

In analogy to the regression models in Table 6, an analyses of covariance is executed. In Table 7 results are reported of a model with the post-test (1988) as the response variable, and the pre-test (1987) as covariate. The factor 'program' has four conditions, Control, RT, NORM and BOTH.

Table 7. Model 1: Ancova with pre- and post-test alcohol

	SS1	F1	Sig of F1
Pre-test	876.13	1456.43	0.000
Program	7.54	4.18	0.006

Program Conditions	Model 1	
Mean Control group	- 0.01	N=671
Mean RT	- 0.06	N=654
Mean NORM	0.10	N=462
Mean BOTH	0.01	N=591

The results in Table 7 show that the factor: Program, with its four categories, has a significant effect ($p=0.006$). The bottom half of the table shows the adjusted means for each of the conditions. Remember that the pre- and post-tests are constructed so that a high positive score is low or no alcohol involvement and high negative is high alcohol involvement, with a mean of zero and a standard deviation of one. The four means show that the overall significant effect for the factor 'Program' is partly due to the negative effect of the drug prevention program RT. RT has, the largest deviation (-0.06) from the overall mean (0.00), making the program with the highest mean alcohol use. The control group has a mean of around zero (0.01), which is equal to the mean of the classes that received NORM and RT together (BOTH). The conclusion based on this analysis is different from the one obtained with multiple regression. In the regression analyses the model fit was not improved by adding the program conditions, while in Ancova the F test shows a significant effect for the factor 'Program'.

Last effect is mainly due to the large difference between RT and NORM, not because of the difference between NORM and the control group.

Since it is still unclear what to think of the program NORM the percentage of abstainers for the four program conditions are calculated and reported in Table 8. Abstainers are defined as the percentage of students that do not drink in 1987, and still don't drink a year later, in 1988. If less students start drinking in one of the programs conditions, as compared to the control group, that drug prevention programs is successful in refraining more students from drinking. The percentages are calculated based on item 19: "How many drinks did you have in your life?". This question is answered in 1987, as well as in 1988. Students who reported that they had 'a sip' to drink are counted as abstainers.

Table 8. Percentage of abstainers for Item 19 in 1988.

	Control*	BOTH*	RT*	NORM*
Item 19	74% (361 out of 489)	76% (342 out of 449)	72% (323 out of 447)	77% (248 out of 324)

Note that the percentages are not abstainers over the total sample, but the percentage abstainers of the original group of abstainers a year earlier. Hence smaller numbers are reported for each group than in Table 7.

The percentages reported in Table 8 show that for all programs including the control group, the number of abstainers are less in 1988 compared to the previous year. If the percentage abstainers in the RT, NORM and BOTH are compared with the control group (74%) it shows again that RT is the least successful condition (72%), while NORM is the most successful (77%). The difference between the RT and NORM is statistically significant, but that is an irrelevant conclusion, since programs need to be compared with the control group. The control group shows a difference of 3% with the NORM program, which is a difference equal to nine students. It is obvious that such a small difference is neither statistically, nor practically significant. More so since the comparison is based on one single item, which has most likely measurement error, as was illustrated in Table 1. The item is also a self report, and "one must always be cautious when interpreting analyses based on a single method of measurement." (Donaldson, Graham and Hansen, 1994, p.212).

Multi Level Analyses with respondents' Pre- and Post-test

Since the data is based on observations nested in existing classes, intraclass correlation may be present. Intraclass correlation effects the standard errors of regression coefficients in a way that leads to an under estimation. As shown in Barcikowski (1981), the presence of an intraclass correlation has an effect on the alpha level of the F-test in analyses of variance, leading to too liberal tests of significance. No significant effect of drug prevention programs is found in the traditional analyses, which makes the expectation that such an effect will not show up in multilevel analysis even smaller, due to stricter tests of significance in methods that do take intraclass correlation in account. Based on these results as well as based on the table with abstainers, no main effects of programs are expected to be found in the multilevel analysis. For reasons of comparison the programs NORM, RT and BOTH are included as explanatory variables in the reported multilevel analyses. For all following multilevel analyses the software ML3 (1989) is used.

Since the previous analyses have indicated that it is most likely that main effects of drug prevention programs will not be found, analyses are used to test if moderator effects of program conditions are present in the data, especially of NORM. Several models test the theory that effect of programs are not equal for all students, but have interaction effects. In the literature it is suggested that interactive effects may exist between drug prevention programs and individual characteristics, although "traditional analyses fail to detect important effects" (see Donaldson et al. 1995, p.5). Multilevel analyses are suited to test interactions, such as the moderator effect of NORM in lowering the strength between pre- and post-test.

Fitting the same model again

The traditional analyses show that drug prevention programs have no significant results other than in the wrong direction. This is verified with a multilevel analysis using the same variables (see the Tables 6, and 7, where pre-test is predicting post-test, together with the three programs NORM, RT, and BOTH. Programs are defined in multilevel analyses as second level explanatory variables, since they are measured at the class level. The control group is set to zero. Results are reported in Table 9.

Table 9. Multilevel analysis with alcohol 1988 as the response variable.

	Model 1		Model 2	
	Par-estimate	st-error	Par-estimate	st-error
intercept	-0.03	0.02	-0.03	0.03
pre-test	0.68	0.03 **	0.68	0.03 **
RT			-0.05	0.04
NORM			0.07	0.04

BOTH			-0.004	0.04
Deviance	5458.21		5451.24	
	Par-est	st- error	Par-est	st-error
Variance level 1	0.56	0.02 **	0.62	0.02 **
Variance level 2				
intercept	0.01	0.01	0.02	0.01 *
slope	0.05	0.01 **	0.05	0.01 **
covariance	-0.03	0.01 *	-0.03	0.01 *

* $p < .05$ ** $p < .01$

Both models in Table 9 have only pre- and post-test as student level explanatory variables, while in the second model (Model 2) the three program conditions are added to the model. The results show that again pre-test predicts the post-test, but program conditions do not add significantly to an explanation of the post-test variation. Several sources in the table support that finding. The individual coefficients of the three conditions, RT, NORM and BOTH, show no significant effects after correction for pre-test (see Model 2 in Table 9). A better way for testing the effects is looking at the model fit. Model fit can be checked by taking the difference in deviance between Model 1 and Model 2, which is 6.93, and compare this difference to the degree of freedom (dfr) lost. Comparing 6.93 with 3 dfr shows that Model 2 does not significantly improvement the fit compared to model 1. The same conclusion is reached here as in the regression analysis (see Table 5a) where the R^2 did not change by adding the three program conditions. There is still a third way to check if adding program conditions improve the model, twchich is by comparing the variance component of the intercept over models. The variation in the intercept in Model 1, the model without the program conditions, is not significant (0.01 with a standard error of the same magnitude). The zero variance of the intercept indicates that no differences in the mean level of post-test alcohol involvement over school classes is present. In other words, there is no potential between school class variation to be explained by any school class level characteristics, including programs. All classes behave in the same way, after correction for pre-test. The only promise in Model 1 is in the significant slope variation of the pre-test, 0.05 with a standard error of 0.01. In the following models I will try to explain this variation among school classes by adding cross-level interactions with one of the program conditions and pre-test.

Multilevel Analyses with respondents pre- and post-test, 'friends' and 'social'

In the following analyses the earlier constructed variables for friends drinking behavior and the behavior in the social environment of the respondents is added. To test the hypothesis that the

most successful program NORM has a lowering effect on the relationship between pre- and post-test, a cross level interaction is introduced in the model (see Interaction Norm*pre-test in Table 10). I expect that the program lowers the strength of the relation between pre- and post-tests. The results are reported in Table 10.

Table 10. Multilevel analysis with alcohol 1988 as the response variable.
Model 1 with main program effect.
Model 2 with cross-level interaction between NORM and alcohol pre-test.

Model 1: Main Effects			Model 2: Cross-level interactions	
	Par-estimate	st- error	Par-estimate	st-error
intercept	-0.03	0.03	-0.04	0.03
alcohol pre-test	0.52	0.03**	0.55	0.03**
friends alc	0.19	0.02**	0.19	0.02**
social alc	0.06	0.02*	0.06	0.02*
RT	-0.05	0.04	-0.04	0.04
NORM	0.07	0.04	0.11	0.05*
BOTH	-0.02	0.04	-0.02	0.04
Interaction Norm * pre-test			-0.10	0.06
Deviance	5350.65		5348.09	
Variance level 1	0.53	0.02**	0.53	0.02**
Variance level 2				
intercept	0.01	0.01	0.01	0.01
alcohol slope	0.05	0.01**	0.04	0.01**
covariance	-0.03	0.01*	-0.03	0.01*

* $p < .05$ ** $p < .01$

At the student level the post-test construct for alcohol is used again as response variable, and pre-test alcohol use, friends' alcohol involvement (friends), and the social context ('social', adults that are drunk, or offer drinks) as explanatory variables. At the school class level drug the three drug prevention programs RT, NORM and BOTH are again included as explanatory variables. In Model 2 of Table 10 the hypothesis is tested that NORM has a cross-level interaction effect with pre-test. If a significant interaction is present, lowering the relationship between pre- and post-test, the hypothesis is supported that drug prevention effects are indirect. Model 2 shows an interaction effect that is in the right direction (negative), but it is not significantly different from zero.

Another way is comparing deviances over a model with and without this cross-level interaction. The deviances over both models in Table 10 show that the fit of Model 2 is not significantly better than Model 1 (the difference in deviance is 2.56 with 1 dfr). A suppressor effect is present in

Model 2, where the addition of the interaction term has changed the coefficient of NORM and its standard error in a way that NORM appears to be significant in Model 2. This significant effect is an artifact of regression models when correlated variables with opposite effects are added to a model. In the example, the interaction coefficient has a sign in the opposite direction of the NORM coefficient, thus enhancing the effect of NORM.

The fit of a model is greatly improved by the addition of the two student level variables 'social' and 'friends'. Comparing model fits (Model 2 in Table 9 compared with with Model 1 in Table 10), an improvement of fit of $5451.24 - 5348.09 = 103.15$ is found. Compared to the loss of 2 dfr, this is by all standards a very significant improvement.

However no evidence is found of drug prevention program effects on post-test alcohol involvement, neither in main effects nor in cross-level interaction effects.

In the next analyses the hypothesis is tested that high mean levels of pre-test alcohol involvement in school classes are interacting with prevention program conditions, as was hypothesized by Hansen and Graham (1991) and Graham et al. (1991).

Multilevel Analysis with Means

Since no longer main effects of drug prevention programs are expected to be present in this data, the analyses proceeds as an exploration of possible interaction effects of drug prevention programs, interactions with special types of students or with mean levels of alcohol use in school classes. In the literature it is suggested that such interactive effects exist between drug prevention programs and individual characteristics, although "traditional analyses fail to detect important effects" (see Donaldson et al. 1994, p.5). Multilevel analysis is designed to test this type of so called 'cross-level'¹ interactions, such as the effect of the most promising drug prevention program NORM on pre-test. The hypothesis is, that NORM will lower the strength of the relationship between pre- and post-test. If such an effect is found NORM has a moderator effect, lowering the effect between individual characteristics.

One of the goals of the normative education curriculum (NORM) in the AAPT study is to demonstrate to students that the actual use among students in the school is much lower than students think or perceive. That is, the common statement, "everyone is doing it" is simply wrong. However, if there happens to be relatively more drinking and other substance use in a particular classroom, the credibility of this normative education message may be seriously

¹ the name 'cross-level interactions' means that a characteristic of the context (the school class, or the special program delivered to that class) interacts with student characteristics. Since students are defined as level 1 units nested within level 2 units (the class), such an interaction is called a 'cross-level' interaction.

undermined. Hansen and Graham (1991) wrote: "it has long been suspected that peer pressure is a major cause of onset of use of common substances" (p.425). Based on this notion the hypothesis is tested that school classes with high alcohol use have lower or no program effects, compared to classes with low average alcohol involvement. This hypothesis is tested by constructing two averages for the amount of alcohol involvement in the class; the class means for pre-test and the mean for friends' alcohol involvement. These two means are interacting with the program NORM in the next model (see the terms 'Interaction Norm' and 'Mean-friends alc Interaction Norm' in Table 11).

Table 11. Multilevel analysis with cross-level interaction with NORM and Mean.

Model 1: Class Mean of respondent			Model 2: Class Mean of Friends	
	Par-estimate	st- error	Par-estimate	st-error
intercept	-0.03	0.03	-0.03	0.03
alcohol pre-test	0.52	0.03**	0.52	0.03**
social alc	0.06	0.02*	0.06	0.02*
Friends alc	0.19	0.02**	0.19	0.02**
RT	-0.05	0.04	-0.05	0.04
NORM	0.07	0.04	0.07	0.04
BOTH	-0.02	0.04	-0.02	0.04
Interaction Norm *				
Mean-friends alc			-0.06	0.10
Interaction Norm *				
Class Mean Resp	0.14	0.11 (n.s.)		
Deviance	5349.22		5350.25	
Variance level 1	0.53	0.02**	0.53	0.02**
Variance level 2				
intercept	0.01	0.01	0.01	0.01
alcohol slope	0.05	0.01**	0.05	0.01**
covariance	-0.03	0.01	-0.03	0.01

* $p < .05$ ** $p < .01$

Note: The deviance of the same model, but without a cross-level interaction is 5350.65

The hypothesis tested in Model 2 of Table 10 is that in classes where alcohol involvement is high, program NORM effects are less. The theory underlying NORM is that overall high use in a school class interacts with the program effectiveness of NORM as measured in the interaction term between NORM and the class 'Mean Friends' in Model 2. If this interaction would be significant and negative, it supports the hypothesis that a higher mean level of alcohol of friends in a class, lowers the positive effect of NORM. Since the coefficient for the interaction term is negative, but

not statistically significant the hypothesis is not supported. The results of both analyses in Table 11 show again that again pre-test, 'social' and 'friends' have highly significant coefficients, while program conditions do not. In both models the interactions between NORM and Mean levels of alcohol involvement are not significant. The results in Table 11 do not add any new information to what we already obtained from Table 10. The same is obvious from the model fit among models. Using the deviance again, with dfs it shows that Model 1 of Table 10, is very close in deviance to the one observed in Table 11, which are respectively 5349.22 and 5350.25. The deviances are not significantly different from each other, which makes the model with the most degrees of freedom the best model that fits the data, which is model I in Table 10.

CONCLUSION

Given the discussion in this paper about the ecological fallacy, it is not surprising that the result of Hansen and Graham's (1991) analyses with the same data set differs from the ones reported here. Based on their class level analysis they report "that a p value of 0.0011 indicates a significant reduction in onset (of alcohol use) attributable to normative education" (l.c. p. 414). The fallacy in last citation is twofold. The analyses were executed at class level, and all that one can conclude based such analyzed is that classes receiving normative education have, on average, a reduction in onset of alcohol use, which may be is attributable to NORM. The next fallacy is that such causal statements are hard to defend when existing classes are used. This will be discussed later. First I want to show that the method used by Hansen and Graham (1991) is a fairly common way to analyze data in drug prevention research (e.g. Dukes, Ullman and Stein, 1995.) One of the reasons is that reviewers have systematically rejected papers based on student level analyses, out of concern that a possible intraclass correlation will distort results. But that results of aggregated analyses are not necessarily the same as the ones obtained from student level analyses is largely ignored by the same reviewers. That such differences can be extensive is illustrated here, as well as in the two papers by Duke, Ullman and Stein (1995 and 1996). They analyzed data at student level (l.c. 1996) finding no program effects, which contradicts their earlier findings (l.c. 1995) based on class level analyses, where such effects were found.

That my results contradict the results reported by Hansen and Graham (1991) with the same data can be explained by the difference in data analyses, but also by the use of a different response variable. Hansen and Graham dichotomized the responses to three items, the life time alcohol use (item 19), alcohol use in past month (item 20) and in past week (item 21). The code one is used to indicate alcohol use (from half a drink up to more than 11 drinks, see appendix for categories of

the same variables), and zero for no use, including a "sip". After this data reduction the student data are aggregated to class level. As a result of different choice made in the data analyses different results are obtained. To evaluate what choices will yield the most reliable results, the reasons for making such a choice need to be clear. My choice of scaling technique and analysis method is based on assumptions regarding the way the data are generated, and are discussed and described in a careful way. Given that the zero effect of NORM found in the multilevel analyses are supported by carefully chosen exploratory analyses the 'successful' program NORM can be said to have neither statistically significant effects, nor important effects. NORM is neither effective as a moderator, nor has it significantly lower number of abstainers. Although the percentage of abstainers from year 87 to 88, calculated over program conditions, shows that NORM has the lowest number of students changing from abstainers to alcohol users, and RT shows the usual high numbers, the difference is neither large enough to be statistically significant, nor important. The same result is obtained in a comparison using item 35 (see appendix), where the question is asked how many times the respondent has been drunk. An increase is observed over the year in the number of students that answer yes to this question (either one, two or more times) among all programs, and is the largest in the control group and the smallest in the NORM group, but again the difference is too small to reach a statistical significance level.

After finding no effects of drug prevention programs I explored the rich data set in search for student risk factors in relation to alcohol use. 'At risk behavior' of students is defined as: "the action of a person or the environment that raises the risk for future alcohol abuse". This concept is used in the drug prevention literature and programs are developed to counter this risk or, at least, lower the risk. My analyses indicate that drinking in the environment of the respondent by friends and adults is related to respondents' alcohol involvement. That finding defines these students as being 'at risk'. The interactions of NORM with friends drinking and respondents drinking are constructed to test if NORM has a moderator, by lowering these relationships. But no moderator effects of NORM are found. If drug prevention programs fail, it may be because the influence of the environment is stronger than the influence of cognitive lectures and exercises, such as offered in drug prevention programs, either in RT, or in NORM. The analyses results support the findings of Kandel (1974) that the example set by parents and peers are a crucial factor in drug use.

Other risk factors are present in this data. In analyses not reported here (see Kreft, 1996) I found that 'trouble' in school, low grades, and 'rebelliousness' (see Kreft, 1997) are factors related to alcohol use. Although the literature mentions risk related to parental behavior, such as a bad or indifferent relationship between parents and respondents' friends, or low degree of parental guidance and/or low degree of parental trust, no such effects are found to be significant in this data (see Kreft 1996). In Kreft (1997) it is reported that if pre-test alcohol involvement is controlled

for in this data, the effects of sex, race and ses are no longer important (see Kreft 1994).

Since factors that cause respondents to get involved with alcohol cannot be determined by this data analyses only, without strong support of a theory, the conclusions drawn from my analysis can only be to exclude not to include. This is even more true for results based on complicated models (see also Draper, 1995). The failure to reproduce the earlier reported effect of the program NORM with the methods used in my analyses removes the support of the hypothesis that NORM is effective. The null hypothesis need to be retained, which is that for this data school based drug prevention programs are not effective, irrespective of the way the message is delivered.

The case is made as strong as it gets in this paper, using homogeneity analyses, where variables are scaled to construct a more reliable and more global scale, while also enhancing the validity of the measurements. The multilevel analyses are executed at the proper level, that of the student, while descriptive statistics are used to underscore their findings.

REFERENCES.

- Aitkin, M.A. and Longford, N. (1986). Statistical modelling in school effectiveness studies. *Journal of the Royal Statistical Society*, 149A, 1-43.
- Bandura, A. (1977). *Social learning Theory*. Englewood Cliffs, N.J.: Prentice Hall.
- Bandura, A. (1986). *Social Foundations of thought and action: A social cognitive theory*. Englewood Cliffs, N.J.: Prentice Hall.
- Barcikowski, R.S. (1981). Statistical Power with Group Mean as the Unit of Analysis. *Journal of Educational Statistics* 6, 3: 267-285.
- Brown, Joel H. and Jordane E. Horowitz (1993). Deviance and Deviants. Why adolescent substance use prevention programs do not work. *Evaluation Review*, 17, 5, 529-555.
- Brown, Joel H. and Marianne D'Emidio Caston (1995). On Becoming "At Risk" Through Drug Education: How symbolic Policies and Their Practices Affect Students. *Evaluation Review*, 19, p. 451-491.
- Benzecri, J.P. (1973). *Analyse des donnees*. Dunod, Paris.
- Burstein, Leigh (1980). Assessing differences between group and individual level regression coefficients: alternative approaches. *Sociological Methods and Research*, 7, 5-27
- Carbonari, Joseph P. Philip W. Wirtz, Larry R. Muenz, and Robert L. Stout (1994). Alternative Analytical Methods for Detecting Matching Effects in Treatment Outcomes. *Journal of Studies on Alcohol*. Supplement #12, 83-90.

Categories (1989) SPSS.

De Leeuw, Jan and Kreft, Ita G.G. (1986). Random Coefficient Models for Multilevel Analysis. *Journal of Educational Statistics*, 11, 1, 57-86.

Draper, David (1995). Inference and Hierarchical Modeling in the Social Sciences. *Journal of Educational and Behavioral Statistics*, 20, 2, 115-148

Donaldson, Stewart I., John W. Graham, and William B. Hansen (1994). Testing the generalizability of intervening mechanism theories: understanding the effects of adolescent drug use prevention interventions. *Journal of Behavioral Medicine*, 17, 2.

Donaldson, Stewart, I, John W. Graham, Andrea, M. Piccinin, and William B. Hansen. (1995). Resistance-skills training and onset of alcohol use: Evidence for beneficial and potential harmful effects in public schools and in private Catholic schools. *Health Psychology*, 14, 291-300.

Dukes, Richard L., Jodie B. Ullman, and Judith A. Stein (1995). An evaluation of D.A.R.E. using a solomon four-group design with latent variables. *Evaluation Review*, 19, 4, 389-409.

Dukes, Richard L., Jodie B. Ullman, and Judith A. Stein (1996). Three-year Follow-Up of Drug Abuse Resistance Education (D.A.R.E.). *Evaluation Review*, 20, 1, 49-66.

Ennett, Susan T., Nancy S. Tobler, Christopher L. Ringwalt, and Robert L. Flewelling (1994). How effective is drug abuse resistance education? A meta analysis of the project DARE outcome evaluations. *American Journal of Public Health*, 84, 9, 1394-1401.

Graham, John, W. , G.S. Marks, and William. B. Hansen, (1991). Social Influence processes affecting adolescent substance use. *Journal of Applied Psychology*, 76, 291-298.

Graham, John, W., Linda M. Collins, Stuart E. Wugalter, N.K. Jacob Chung, and B. Hansen (1991). Modeling transitions in latent stage-sequential processes: A substance use prevention example. *Journal of Consulting and Clinical Psychology*, 59, 1, 48-57.

Greenacre, M.J. (1984). *Theory and Applications of correspondence analysis*. Academic Press, London.

Guttman, Louis A. (1941). The quantification of a class of attributes: a theory and a method of scale construction. in: P. Horst (ed.), *The Prediction of Personal Adjustment*., New York, SSRC.

Hannan, M.T. (1974). Problems of aggregation. In: H.M. Blalock, (ed.) *Causal models in the social sciences*. Chapter 27, 473-508. Chicago, Aldine

Hansen, W. B. (1993). School-Based Alcohol Prevention programs. *Alcohol Health & research World*, 17, 1, 54-60.

Hansen, W.B. and J.W. Graham (1991). Preventing Alcohol, Marijuana, and Cigarette Use among Adolescents: One-year result results of the Adolescent Alcohol Prevention Trial. *Preventive Medicine*, 20, 414-430.

Hedeker, D., R.D. Gibbons, and J.M. Davis (1991). Random regression models for multicenter

- clinical trials data. *Psychopharmacology Bulletin*, 27, 73-77.
- Hedeker, Donald, Robert D. Gibbons and Brian R. Flay (1995) Random-effects regression Models for Clustered Data: With an example from Smoking Prevention Research. *Journal of Consulting and Clinical Psychiatry* (in press).
- HOMALS (1989). *SPSS-Categories*. Chicago: Author.
- Jessor, R. and Jessor, S.L. (1977). *Problem behavior and psychosocial development*. New York: Academic Press.
- Kandel, D.B.(1974). Interpersonal influences on adolescent illegal drug use. In: *Drug use: Epidemiological and sociological approaches*. E.Josephson and E.E. Carroll (Eds.), 207-239. Washington, DC: Hemisphere.
- Kreft, Ita G.G. and Edith D. De Leeuw (1988). The Seesaw effect: a multilevel problem? *Quality and Quantity*, 22, 127-137.
- Kreft, Ita G.G. (1994) Multilevel Models for Hierarchically Nested Data: Potential Applications in Substance Abuse Prevention Research. Pp. 140-184. In: *Advances in Data Analysis for Prevention Intervention Research*. Linda M. Collins and Larry A. Seitz (Eds.), NIDA Monograph #142, National Institutes of Health, Rockville, MD
- Kreft, Ita, G.G. (1996). The Evaluation of the Effects of drug Prevention Programs on Junior High School Students' Alcohol Consumption with Homogeneity and Multilevel Analyses. PERC Report #1, Charter School of Education, California State University, Los Angeles
- Kreft, Ita G.G. (1997) The interactive effect of alcohol prevention programs in high school classes. In: *New methodological developments in prevention: Alcohol and Substance Abuse Research*. Bryant, Kendall, J., Michael Windle, and Stephen West (Eds.). APA Publication
- Kreft, Ita G.G., Jan de Leeuw, and Rien van der Leeden (1994). Review of five multilevel analysis programs: BMDP-5V, GENMOD, HLM, ML3, VARCL. *The American Statistician*, 48, 4, p. 324-335.
- Laird, N.W., and Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38, 963-974.
- Longford, N.T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, 74, 817-827.
- May, Carl (1993). Resistance to peer group pressure: an inadequate basis for alcohol education. *Health Education Research: Theory and practice*, 8, 2, 159-165
- Meulman, Jacqueline (1982). *Homogeneity Analysis of Incomplete Data*. Leiden, DSWO Press.
- ML3 (1989). Software for two- or three-level analysis written by Rabash. London: Institute of Education, University of London
- Robinson, W.S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15, 351-357.

- Van de Geer, John P. (1993a). *Multivariate Analysis of Categorical Data: Theory*. Advanced Quantitative Techniques in the Social Sciences Series, #2, Sage, Newbury Park.
- Van de Geer, John P. (1993b). *Multivariate Analysis of Categorical Data: Applications*. Advanced Quantitative Techniques in the Social Sciences Series, #1, Sage, Newbury Park.
- Wang C. (1993). *Sense and Nonsense of Statistical Inference*, New York, Marcel Dekker
- Raudenbush, S. and Bryk, A.S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 59, 1-17.
- Robinson, W.S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15, 351-357.

APPENDIX
HOMALS CATEGORY QUANTIFICATION TABLES

Item 19: How many drinks of alcohol have you had in your whole life?

Pre-test	Post-test	Original code and Wording	N 1987	N 1988
0.64	0.70	1 none	1002	743
0.42	0.60	2 only sip (for religion)	355	304
0.18	0.36	3 only a sip (not religion)	760	756
-0.25	0.05	4 part or all of a drink	226	240
-0.52	-0.21	5 2 to 4	227	280
-0.97	-0.56	6 5 to 10	159	224
-1.59	-1.02	7 11 to 20	129	178
-2.23	-1.62	8 21 to 100	103	200
-3.07	-2.75	9 more than 100	50	105
-0.31	-0.74	Missing	16	17

Item 20: How many drinks of alcohol have you had in the past month?

Pre-test	Post-test	Original code and Wording	N 1987	N 1988
0.35	0.44	1 none	2383	2209
0.09	0.32	2 only sip (for religion)	130	100
-0.90	-0.58	3 only a sip (not religion)	232	243
-1.70	-1.06	4 part or all of a drink	99	159
-2.25	-1.67	5 2 to 4	115	161
-3.18	-2.18	6 5 to 10	35	93
-4.13	-2.83	7 11 to 20	13	36
-4.07	-3.63	8 more than 20	13	36
0.00	0.00	Missing	7	10

Item 23: How many drinks of alcohol have you had in the past week?

Pre-test	Post-test	Original code and Wording	N 1987	N 1988
0.23	0.27	1 none	2676	2610
-0.19	0.05	2 only sip (for religion)	67	63
-1.46	-1.20	3 only a sip (not religion)	116	96
-2.15	-1.72	4 1/2 or less	50	87
-2.44	-1.83	5 1	50	66
-3.35	-2.48	6 2 to 4	35	60
-3.55	-2.95	7 5 to 10	12	31
-4.24	-3.82	8 11 or more	11	22
0.00	0.00	9 missing	10	12

Item 24: How many days in the past month have you had alcohol to drink?

Pre-test	Post-test	Original code and Wording	N 1987	N 1988
0.32	0.41	1 none	2586	2402
-1.32	-0.97	2 1	240	286
-2.27	-1.72	3 2 to 3	117	198
-3.06	-2.17	4 4 to 7	48	90
-4.08	-3.09	5 8 to 14	14	23
-3.32	-3.43	6 15 to 30	12	29
0.00	0.00	7 missing	10	19

Item 25: How long has it been since you had any alcohol to drink?

Pre-test	Post-test	Original code and Wording	N 1987	N 1988
-1.88	-2.12	1 Less than 24 hours	47	55
-2.25	-1.95	2 > a day, < a week	147	197
-1.39	-1.14	3 > a week, < a month	269	386
-0.36	-0.13	4 > a month, < 6 months	438	524
0.09	0.21	5. > 6 months, < a year	302	282
0.29	0.45	6. > a year	567	604
0.61	0.69	7. I never had any alcohol	1243	977
0.00	0.00	8 Missing	14	22

Item 28: Think of the day during the past month when you drank the most alcohol.
How many drinks did you have that day?

Pre-test	Post-test	Original code and Wording	N 1987	N 1988
0.58	0.65	1 I never drink	1459	1256
0.05	0.22	2 No alcohol past month	766	863
-0.41	-0.29	3 sips	384	328
-1.10	-0.94	4 1	160	210
-1.81	-1.40	5 2	82	109
-2.02	-1.68	6 3	51	77
-2.15	-2.04	7 4	31	45
-2.78	-2.36	8 5 or more	81	140
0.00	0.00	9 missing	13	19

Item 29: How often do you imagine yourself having a drink of alcohol?

Pre-test	Post-test	Original code and Wording	N 1987	N 1988
-1.88	-2.18	1 often	59	77
-1.61	-1.47	2 sometimes	232	337
-0.39	-0.18	3 hardly ever	819	889
0.42	0.48	4 never	1905	1726
		5 Missing	12	18

Item 32: Do you think you will drink alcohol in the next couple of months?

Pre-test	Post-test	Original code and Wording	N 1987	N 1988
-2.79	-2.17	1 Yes	117	223
-1.37	-1.05	2 Probably	303	403
-0.38	-0.14	3 I don't think so	536	639
0.46	0.57	4 No	2067	1756
		Missing	4	26

Item 33: Do you think you will ever drink alcohol every day?

Pre-test	Post-test	Original code and Wording	N 1987	N 1988
-1.55	-2.19	1 Yes	28	37
-1.78	-1.25	2 Probably	75	71
-0.93	-0.92	3 I don't think so	349	418
0.19	0.22	4 No	2566	2496
		Missing	9	25

Item 34: Do you think you will ever drink alcohol every month?

Pre-test	Post-test	Original code and Wording	N 1987	N 1988
0.39	0.47	1 No	2195	1915
-0.55	-0.32	2 I don't think so	452	585
-1.23	-1.17	3 Probably	293	381
-2.03	-1.98	4 Yes	114	140
		Missing	9	26

Item 35: How many times have you ever been drunk?

Pre-test	Post-test	Original code and Wording	N 1987	N 1988
0.28	0.35	1 Never	2520	2340
-0.72	-0.47	2 only once	287	298
-1.94	-1.31	3 2 to 4 times	140	253
-2.42	-2.17	4 5 to 10 times	43	75
-3.43	-2.82	5 11 to 20 times	14	29
-4.34	-3.52	6 > 20 times	15	28
0.00	0.00	7 Missing	8	24

Item 38: Do you think you will get drunk in the next couple of months?

Pre-test	Post-test	Original code and Wording	N 1987	N 1988
0.23	0.33	1 No	2676	2464
-1.29	-0.96	2 I don't think so	211	348
-2.42	-1.94	3 Probably	85	136
-3.14	-3.11	4 Yes	41	71
		Missing	14	28

ⁱ Nida grant # DA09649-02 provided me with a total of six month over a period of two years (1994 -1996) to reanalyze the Adolescent Alcohol Prevention Trial data, collected by Hansen and Graham. The data are made available by John W. Graham

ⁱⁱ D.A.R.E. is the copyrighted acronym for Drug Abuse Resistance Education. It is administered by D.A.R.E. a nonprofit organization based in Los Angeles, California. The program is administered starting in the last grades of elementary school. Police officers are trained to teach students to resists drug offers, and instead accept a drug free live style. This program resembles the earlier mentioned RT program

ⁱⁱⁱ See Stewart I. Donaldson, John W. Graham, Andrea M. Piccinin, William B. Hansen. Resistance Skills training and Alcohol Use Onset: Evidence fro Beneficial and Potential Harmful Effects in Public and Private Catholic Schools, *Health Psychology*

^{iv} RT is an acronym for resistance training. It is designed to help kids see the kinds of pressure to use drugs by teaching them skills to resists such pressure without losing friends. The program is based on the assumption that kids want to resists drug offers, but simply lack the proper skills.